Evaluating Student Models

Adaeze Nwaigwe
University of Maryland University College
3501 Unversity Blvd East
Adelphi, MD 207831 412 608 8747
adaeze.nwaigwe@faculty.umuc.edu

ABSTRACT
We use the Additive Factors Model to drive the evaluation of the student model of an Intelligent Tutoring System. Using data from the Andes Physics Tutor, applying the simple location heuristic and implementing the Additive Factors Model tool in the Pittsburgh's Science of Learning Center's DataShop, we discover possible ways to improve the student model of the Andes Intelligent Tutor.

Keywords
Student modeling, learning curves, additive factors model.

1. INTRODUCTION
The quality of student models drive many of the instructional decisions that automated tutoring systems make, whether it is what feedback to provide, when and how to sequence topics and problems in a curriculum, how to adapt pacing to the needs of students and even what problems and instructional materials are necessary [1]. We used the Additive Factors Model (AFM) tool in the Pittsburgh's Science of Learning Center's (PSLC) DataShop to identify areas for improvement in the curriculum for the ANDES Intelligent Tutoring System.

1.1 BACKGROUND
Learning curves derived from student models drive evaluation, revision and improvement of the Intelligent Tutor. The AFM is a statistical algorithm which models learning and performance by using logistical regression performed over the "error rate" learning curve data [1]. If a student is learning the knowledge component (KC) or skill being measured, the learning curve is expected to follow a so-called "power law of practice" [2]. If such a curve exists, it presents evidence that the student is learning the skill being measured or conversely, that the skill represents what the student is learning.

While use of learning curves is now a standard technique for assessing the cognitive models of Intelligent Tutors, the technique requires that a method is instated for attributing blame to skills or KCs. This simply means that each error a student makes must be blamed on a skill or set of skills. Four different heuristics for error attribution have been proposed and tested. These heuristics are guided by whether the method is driven by location – the simple location heuristic (LH), the model-based location heuristic (MLH); or by the temporal order of events – the temporal heuristic (TH), the model-based temporal heuristic (MTH); and whether the choice of the student model is leveraged (MLH, MTH) [3].

2 EVALUATING THE STUDENT MODEL

2.1Adapting the Andes Log data for the AFM Algorithm
The log data used for this work was obtained from the Andes Intelligent Tutor [4] and encompassed four problems in the area of electric field, across 102 students. The data was collected in Spring 2005 at the US Naval Academy during its regular physics class and as part of the PSLC's LearnLab facility that provides researchers, access to run experiments in or perform secondary analyzes of data collected from one of seven available technology-enhanced courses running at multiple high school and college sites (see http://learnlab.org).

Prior to using the AFM tool on the dataset, the simple location heuristic (LH) was applied to error transactions in the Andes log data which had missing KCs. That is, when the Andes failed to assign blame to a KC on an error transaction, the LH will select the first correctly implanted KC in the same location as the error. The LH was applied to about 44% of the original data. Table 1 depicts a summary of the LH data.

2.2 Generating Model Values using AFM
The Datashop's AFM algorithm was used to compute statistical measures of goodness of fit for the model - Akaike Information Criterion (AIC) and Bayesian Information criterion (BIC), as well as to generate learning curves for the Andes log data.

3 RESULTS AND DISCUSSION
We found that there were 5 groups of KCs – "Low and Flat", "No learning", "Still high", "Too Little data" and "Good". The "Low and Flat" group indicated KCs where students likely received too much practice. It appears that although students mastered the KCs they continued to receive tasks for them. It may be better to reduce the required number of tasks or change Andes' knowledge tracing parameters so that students get fewer opportunities with these KCs. The "Still high" group suggests KCs, which students continued to struggle with. Increasing opportunities for practice for these KCs might improve the student model. The "No learning" group indicated KCs where the slope of the predicted learning curve showed no apparent learning. A step towards improving the student model could be to explore whether each of these KCs can be split into multiple KCs. The new KCs may better reflect the variation in difficulty and transfer of learning that may be happening across problem steps, which are currently labeled by each KC. The KCs in the "Too Little data" group seem to be KCs for which students were exposed to insufficient practice opportunities for the data to be meaningful. For these KCs, adding

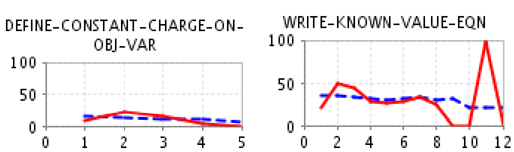
more tasks or merging similar KCs might provide data that is interpretable. The KCs that appeared "Good" may reflect those in which there was substantial student learning. Table 2 shows the different group of KCs, their frequencies and AIC and BIC scores. Figures 1a – 1d show the different groups of KCs. Intercept (logit) and intercept (probability) both indicate KC difficulty. Higher intercept values indicate more difficult KCs. The slope parameter indicates the KC learning rate. Higher values suggest students will learn such KCs faster.

Table 1. LH Data Summary

Number of Students	102
Number of Unique Steps	125
Total Number of Steps	5,857
Total Number of Transactions	71,300
Total Student Hours	107.02
# of Knowledge Component Model	34

Table 2. KC Groups and Statistical Scores

Low and Flat	No Learning	Still High	Too Little data	Good
2	2	4	24	2
# of Knowledge Components				34
AIC				6532.75
BIC				7668.14



KC Name	Intercept (logit)	Intercept (probability)	Slope
define-constant-charge-on-obj-var	1.77	0.85	0.120
write-known-value-eqn	0.63	0.65	0.037

Actual data   predicted

Figure 1a – "Good"

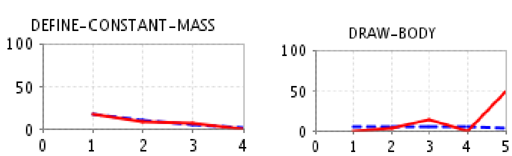
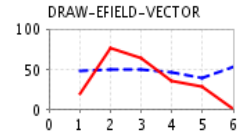
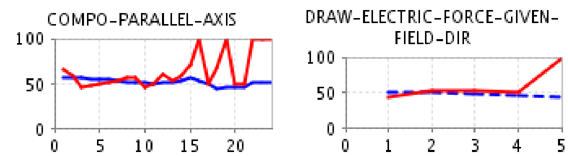


Figure 1b – "Low and Flat"



KC Name	Intercept (logit)	Intercept (probability)	Slope
draw-efield-vector	0.06	0.52	0.000

Figure 1c – "No Learning"



KC Name	Intercept (logit)	Intercept (probability)	Slope
compo-parallel-axis	-0.28	0.43	0.000
draw-electric-force-given-field-dir	-0.01	0.50	0.000

Figure 1d – "Still High"

4 CONCLUSION AND FUTURE WORK

This paper presented how the AFM can be used to evaluate the student model of the Andes Physics Tutor. Refining four of the five groups of KCs identified, might improve the Andes student model. A further approach would to use Learning Factors Analysis [1] algorithm to automatically find better student models by searching through a space of KC models. The next step is to explore these options and measure their effect.

5 ACKNOWLEDGMENTS

Our thanks to the Pittsburgh Science of Learning Center for providing the analysis tool for this work, to Bob Hausmann and Kurt VanLehn for dataset access.

6 REFERENCES

[1] Koedinger, K.R., McLaughlin, E.A., Stamper, J.C. 2012 Automated Student Model Improvement. Proceedings of the 5th International Conference on Educational Data Mining, Chania, Greece, pp. 17–24.

[2] Mathan S. & Koedinger K. 2005. Fostering the Intelligent Novice: Learning From Errors With Metacognitive Tutoring. Educational Psychologist. 40(4), pps. 257–265.

[3] Nwaigwe, A. & Koedinger, K.R. 2011. The Simple Location Heuristic is Better at Predicting Students' Changes in Error Rate Over Time Compared to the Simple Temporal Heuristic. Proceedings of the 4th International Conference on Educational Data Mining. Eindhoven, Netherlands.

[4] VanLehn, K., Lynch, C., Schultz, K., Shapiro, J. A., Shelby, R. H., Taylor, L., et al. 2005. The Andes physics tutoring system: Lessons learned. International Journal of Artificial Intelligence and Education, 15(3), 147-204.